計測自動制御学会東北支部第289回研究集会(2014.6.24)
資料番号289-6

強化学習を用いた6脚ロボットの旋回動作獲得

Acquisition of Turning Motion for a Six-legged Robot Using Reinforcement Learning

○阿部太郎*，釜谷博行*
○Taro Abe*，Hiroyuki Kamaya*
*八戸工業高等専門学校
*Hachinohe National College of Technology

キーワード: 組み込みシステム(Embedded System)，6脚ロボット (Six-legged Robot)，
強化学習 (Reinforcement Learning)

連絡先: 〒039-1192 八戸市田面木字上野平 16-1 八戸工業高等専門学校　電気情報工学科
釜谷博行，Tel.: 0178-27-7283，E-mail: kamaya-e@hachinohe-ct.ac.jp

1. はじめに

近年、ロボットは医療、探査、災害救助など、多種多様な分野で用いられている。中でも、2011 年に起きた東日本大震災の現場においては、さまざまなロボットが、救助活動、がれき撤去、除染などに使用された。しかし、依然このような活動は人力で行われることが多く、二次災害などの危険が伴う。特に原子力発電所での事故処理においては人体への健康被害が著しいものとなる。

このような災害現場で注目されているのが、歩行型ロボットである。歩行型ロボットは平地や荒れ地ともに移動することができ、特に荒れ地においては重心位置を一定に保つことができるため、より安定した移動が可能であるというメリットがある[1]。

しかし、歩行型ロボットにはロボットの関節が増えるほど歩行モーションの作成が困難となる、また、実際に現場において動作させる場合に、ロボットの動作が固定されていると、ロボットは決められた動作しかできないため、歩行ロボットのメリットである荒れ地における安定した歩行性能が十分に発揮されないといった問題がある。そのため、学習を通じて行動の獲得や改善を図る方法として強化学習[2]が提案されており、その代表的なアルゴリズムとしてQ-学習などが知られている。

本研究では、これまで実験用に開発した6脚歩行ロボットを用いて、直進動作の獲得に成功した。直進距離の測定には

PSDセンサを使用し、測定した距離を報酬値として利用した。今回は、旋回動作をQ-学習により獲得させることを目的とする。

2. 強化学習

2.1 概要

強化学習とは未知なる環境下においてエージェント(学習主体)が試行錯誤を通じて得られる報酬をもとに、より良い行動を学習によって獲得する手法である。学習の流れとして、(1)エージェントは環境から状態s_tを知覚する。(2)エージェントは状態s_tに基づいた行動a_tを起こす。(3)起こした行動a_tによりエージェントは状態s_tから新たな状態s_{t+1}へ遷移する。(4)遷移する過程でエージェントは報酬r_{t+1}を得て、この報酬をもとにより良い動作を学習する。以後、$s_t \leftarrow s_{t+1}$として(2)~(4)のサイクルを繰り返す(Fig.1)。

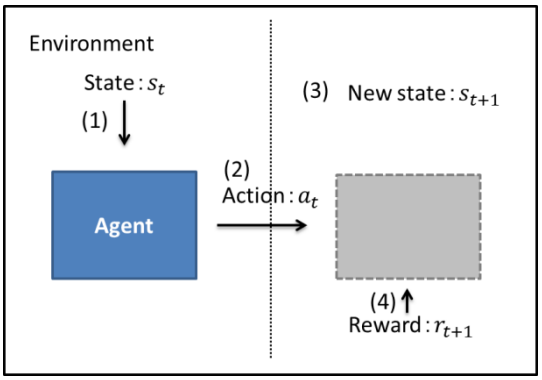


Fig.1 Outline of reinforcement learning

2.2 Q-学習(Q-learning)

Q-学習は方策オフ型のTD学習の一種である。Q学習の利点として、環境がMDP(マルコフ決定過程)であれば最適な方策を獲得できることが保証されている、アルゴリズムが簡潔であるということが挙げられる。

Q-学習は状態行動対に対する行動価値関数$Q(s_t, a_t)$を更新することにより、良好な振る舞いを学習する。Q-学習におけるQ値の更新は次式によりなされる。

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha\left\{r_{t+1} + \gamma \max_b Q(s_{t+1}, b) - Q(s_t, a_t)\right\} \quad (1)$$

$\max_b Q(s_{t+1}, b)$ は遷移した状態s_{t+1}において最も高いQ値を表す。r_{t+1}は行動によって得られる報酬である。また、αは学習率$(0 < \alpha \leq 1)$、γは割引率$(0 \leq \gamma \leq 1)$と呼ばれるパラメータである。これらのパラメータは学習の収束速度と正確さに大きく関わってくるので、適切に設定する必要がある。

2.3 ε-greedy法

Q-学習における行動選択法にはε-greedy法を用いた。ε-greedy法とは1-εの確率でQ値が最も高い行動を、εの確率でランダムな行動をとる行動選択手法である。エージェントは時々ランダムな行動を取ることにより、今まで学習した行動よりもより良い行動を探索することが期待できる。

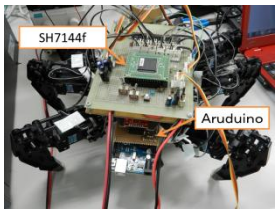
3. システム構成

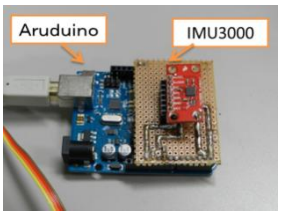
システム構成をFig.2 に示す。システムはロボットとPCから構成されている。ロボットは、近藤科学社のKMR-P4を、脚の数を4脚から6脚へと拡張したものである。KMR-P4に搭載されているマイコンではプログラムによる複雑な制御が不可能で

あったため、SH7144f マイコン(ルネサスエレクトロニクス社、動作クロック48MHz、ROM256KB、RAM8KB)を用いたコントローラを新たに設計した。1脚当たりに搭載されるサーボモータは2個で、ロボット全体では計12個のサーボモータを制御する。

なお、マイコン側のRAM容量が小さいため学習の処理はPC側で行い、PCとロボット間の通信はシリアル通信により行っている。

一方、旋回動作の学習にあたって旋回角度を報酬値として用いるため、角度計測用にジャイロセンサを取り付けた。ジャイロセンサの角度算出の処理は別のマイコンArduino UNO(Arduino Software)で行う。





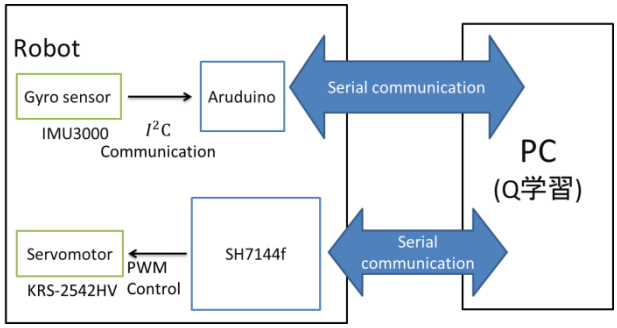


Fig.2　Construction of the system

る数値積分を用いて旋回角度を算出する。まず、角速度$\omega_{(t)}$を一定時間間隔t_0でサンプリングし、取得する。このサンプリング間隔はマイコンのタイマ割り込みにより実現し、$t_0 = 10[ms]$とした。そして、サンプリング前後の角速度値を用いて、次式より角度変化量$\Delta\theta_{(t)}$を算出する。

$$\Delta\theta_{(t)} = \frac{1}{2}\left(\omega_{(t)} + \omega_{(t-1)}\right)t_0 \quad (2)$$

$$\theta_{(t)} = \theta_{(t-1)} + \Delta\theta_{(t)} \quad (3)$$

この計算を繰り返すことにより角度を算出する。

ジャイロセンサによる旋回角度計測の精度について実験する。自作した実験装置をFig.3に示す。ジャイロセンサをこの装置の回転部分に取り付け、回転部分を0(度)~90(度)まで5(度)刻みに動かし、各状態における角度を算出した。ジャイロセンサは小型圧電振動ジャイロモジュール(秋月電子通商)とIMU3000搭載のデジタルIMU(sparkfun)の2種類を試した。

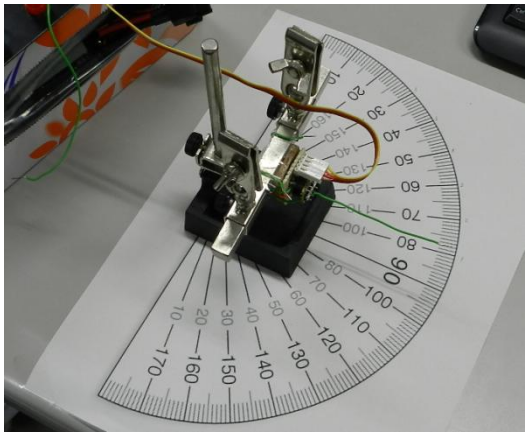


Fig.3　Angle measuring device

4. 旋回角度計測

本研究では旋回角度を計測するためにジャイロセンサを用いた。ジャイロセンサは角速度を出力するため、台形法によ

小型圧電振動ジャイロモジュールは内部に2つのジャイロセンサENC-03R(村田製作所)が搭載されており2軸の角速度を計測することができる。今回は2軸のうち

1軸を使用した。また、安価かつ小型であり、ロボットなどへの搭載が容易である。

このジャイロモジュール用いて角度を計測したところ、正確な角度を算出することができなかった。この原因を調べたところ、センサの出力部に微分回路が設けられており、角速度を正確に出力できないことが分かった。また、ENC-03Rの出力も温度変化の影響を受けやすいため、出力的に安定していないことが分かった。

IMU3000 搭載のデジタル IMU は 3 軸の角速度を計測することができる。マイコンとの通信方式はI^2C通信(400([kHz])で、回路内部にプルアップ抵抗が内蔵されているので、新たにプルアップ抵抗を設置する必要がなく、直接マイコンと接続して使用することができる。また、検出範囲も初期設定より選ぶことができ、用途に応じて使い分けることができる。今回は 3 軸のうち 1 軸のみを使用する。また、ロボットの旋回速度を考慮し、検出範囲を±500[deg/sec]に設定した。

実験結果を Fig.4 に示す。横軸は Fig.3 に示す装置より設定した角度で、縦軸はジャイロセンサにより算出した角度である。横軸と縦軸の値はほぼ一致しており、線形の関係となった。また、90 度旋回時の誤差は-0.54[度]と十分な精度があることを確認した。

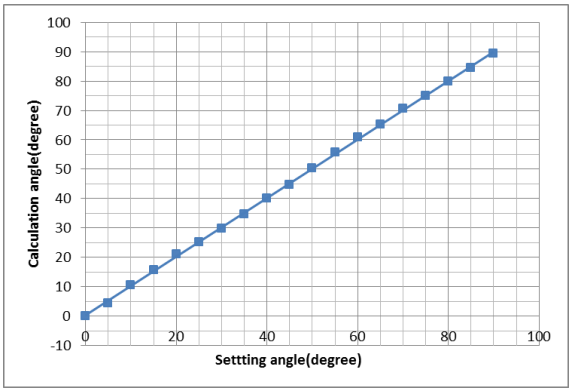


Fig.4　Correspondence of the measured angle and the calculated angle

5.　実験

5.1　パラメータの設定

本研究で使用するサーボモータは-135~135 度の範囲で動作させることが可能であるが、その可動範囲をすべて使うと、脚同士が干渉してしまうので、サーボモータの動作範囲を-15~15 度の範囲に制限する。また、ロボットの状態数の削減のためにサーボモータのとりうる角度を 15、0、-15 度に制限している。

以上のサーボモータの状態によってロボットの脚は、上下3状態、左右3状態の計9状態となる。各状態における行動は上下左右の4つの方向を取ることができる(Fig.5(a))。そして、Fig.5(b)に示すように3脚を1組として同時に動かす。2組あるのでロボット全体の状態数は9×9=81状態、各状態における行動の組み合わせは4×4=16通りとなる。1ステップ毎に各組の脚を同時に動作させる。

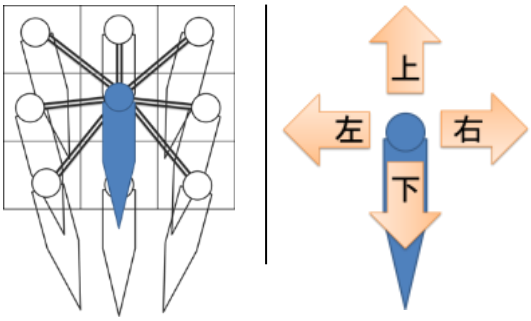
ロボットの単位ステップ当たりの旋回角度の最大化を目標とし、ロボットが毎ステップあたりに旋回した角度に0.1を乗じた値を報酬とした。また、旋回角度が

±2度よりも小さい場合は誤差とみなし、与える報酬値を0とする。なお、隣接した脚同士が干渉しないようにするため、サーボモータを上下左右9状態の範囲外に動作させるような行動を選択したら、負の報酬-5を与える。ロボットが足を引きずるような行動を実行したら、サーボモータに負担をかける行動となるので負の報酬-3を与える。また、収束の速さと安定性を考慮してε-greedy 法における ε の初期値ε_0を0.3と定めている。学習の経過にしたがって ε の値は次式により減少させる。

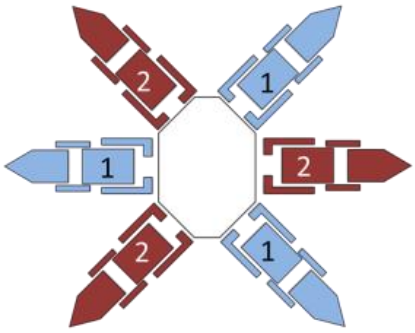
$$\varepsilon = \varepsilon_0(1 - \mathrm{trials}/20) \qquad (4)$$

$(trials$: ステップ数を60で割った商の値$)$

学習率と割引率は予備実験の結果、学習が最も速く収束した値を採用し、それぞれα=0.5、γ=0.5とした。



(a) 9 states and 4 actions



(b)Two sets of leg

Fig.5　State and action of the robot

5.2　旋回動作の獲得

5.1のパラメータを用いて旋回動作の獲得実験を5回行った。結果をFig.6に示す。横軸はステップ数で、縦軸は旋回角度である。旋回角度の計測には外部カメラを用いた。

どの実験においてもステップを重ねるごとに旋回角度が増加しており、また、その傾き(旋回速度)も増加していることがわかる。このことからロボットはより良好な旋回動作を学習によって獲得しようとしていることわかる。旋回動作の獲得には約600ステップ、時間で約30分かかった。

600ステップ以降において、若干グラフの傾きに乱れが見られるが、これは ε-greedy法における行動選択において、 ε の値にしたがって時々ランダムな行動を選択しているためである。しかし、600ステップ経過時においてすでにロボットは良好な旋回動作を学習しているので、すぐにランダムな動作から復帰する。その後、ロボットは再び良好な旋回動作を始め、旋回角度は増加し続けていることが分かる。

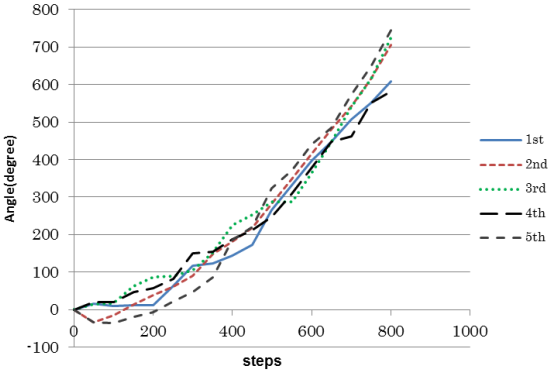


Fig.6　Acquisition of Turning Motion

5.3　獲得した旋回動作の評価

5.2の学習実験において、5つの試行から、それぞれパターンが異なる5つの旋回動作が得られた。ここでは、ロボットが90(度)旋回するのにかかる時間を計測することで得られた旋回動作の評価を行う。また、本評価では比較のためにロボットの状態に基づいて、設計者が考案した自作旋回パターンについても評価を行う。

実験結果をFig.7に示す。各試行で得られた旋回パターンによる旋回時間は一定ではないことがわかる。5番目の学習のように自作旋回パターンよりも速い旋回動作が得られる場合もあるが、全体的には、自作旋回パターンの時間とほぼ同じか、若干時間のかかる旋回動作が得られた。

このように学習結果に違いが生じた原因は、学習終了時の800ステップの時点においてεの値が0になっていない、つまり学習が収束していないことも原因のひとつと考えられる。このことから、ロボットは安定して最適な旋回動作を学習していないといえる。

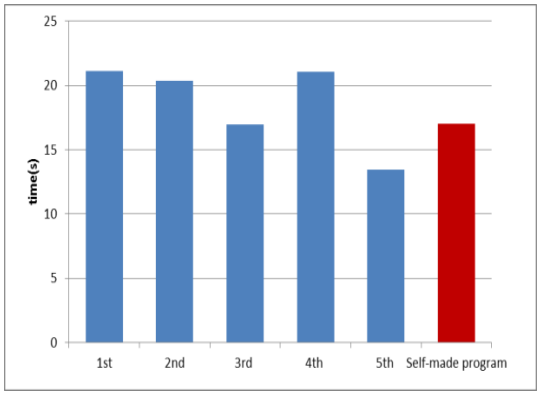


Fig.7　Comparison of turning motion by self-made program and turning motion by learning

6.　まとめと今後の課題

強化学習を適用することにより旋回動作を獲得することができた。しかし、それぞれ獲得した旋回パターンは同一ではなく、最適な旋回動作を安定して学習するには至らなかった。

今後の課題としては、安定して最適な旋回動作を獲得すること、学習を高速化することなどが挙げられる。また、本研究では平坦な床での動作獲得のみであったため、荒れ地等で学習により歩行動作を獲得することも挙げられる。

7. 謝辞

本研究にあたりロボットに関する仕様などをまとめてくださった西佳一郎氏に感謝致します。

参考文献

(1)西澤保輝, 王碩玉：移動形態変更可能なロボットの開発, 日本ロボット学会講演概要集, CD-ROM 3K22 (2007)

(2)三上貞芳, 皆川雅章:強化学習, 森北出版株式会社 (2000)